The second exam covers through lecture 15.2, except that there are not specifically questions about texture synthesis, intelligent scissors, or graph cuts. The exam is roughly split between the fundamentals introduced in the first half of the semester (particularly filtering and template matching; light, color, and contrast enhancement; warping and morphing; single view geometry and camera models) and the second half (image-based lighting; keypoint matching, alignment, and stitching; object/image retrieval and other applications). In addition to these sample questions, the lecture “take home” and online quiz questions can also be used to test your knowledge. I’ve also outlined what I consider to be the key ideas of each section, and nearly all of the exam questions are related to these key ideas.

Filtering

Key ideas

- Design a linear filter based on an equation for local neighborhoods, or to perform an operation such as local averaging, or to achieve a desired effect such as smoothing or edge detection
- Compute the response to a linear filter
- Familiarity with definition and use of padding and kernel size
- Method and appropriate use for non-linear filters, such as median and bilateral filters
- How to perform filtering in frequency domain and representation of images and filters in terms of frequencies

1. Design a 3x3 linear filter, such that the output will be
   $im_{out}(i,j) = 4 \cdot im_{in}(i,j) -\cdot\, im_{in}(i-1,j) - im_{in}(i+1,j) - im_{in}(i,j-1) - im_{in}(i,j+1)$
   where (i,j) is a pixel coordinate that is not on the border, and im_{in} and im_{out} are the input image and output of the filter.

Filter

0	-1	0
-1	4	-1
0	-1	0

2. *W*hich statement about the filter below is most true?

0	0	0
-1	0	1
0	0	0

a. *The filter is a high-pass filter (mostly high frequencies are preserved)*
b. *The filter is a low-pass filter (mostly low frequencies are preserved)*
c. *The filter preserves both high frequencies and low frequencies about equally*
d. *Which frequencies are preserved depends on the input*

A

Filter

-1	0	1
-2	0	2
-1	0	1

*

Input Image

0	6	0	0
0	1	9	0
9	7	1	2
1	9	3	6

➔

Filter Response

x	x	x	x
x			x
x			x
x	x	x	x

3. What is the response at the circled position?
   a. 19
   b. 10
   c. 1
   d. None of these

B

Filter

-1	0	1
-2	0	2
-1	0	1

*

Input Image

0	6	0	0
0	1	9	0
9	7	1	2
1	9	3	6

➔

Filter Response

x	x	x	x
x			x
x			x
x	x	x	x

4. What is the response at the circled position?
    a. -18
    b. 16
    c. -12
    d. None of these

D (-13)

Filter

-1	0	1
-2	0	2
-1	0	1

*

Input Image

0	6	0	0
0	1	9	0
9	7	1	2
1	9	3	6

➔

Filter Response

x	x	x	x
x			x
x			x
x	x	x	x

5. What is the response at the circled position?
    a. 3
    b. 23
    c. -5
    d. None of these

C

6. Match and answer

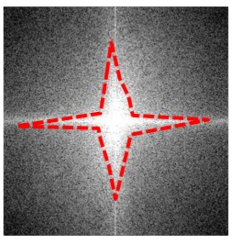
a

b

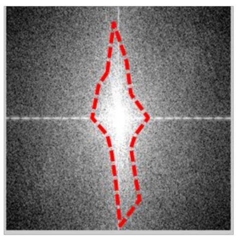
c

d

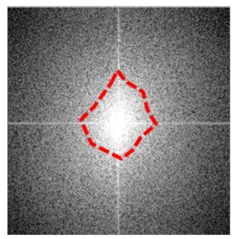
e



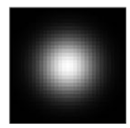
f



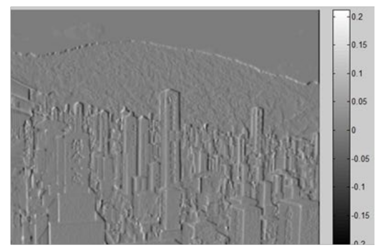
g



h



i



FFT Images (center is 0 frequency)

i. Which FFT image best matches (a)? ___f______

ii. Which FFT image best matches (b)? ___g______

iii. Object __c____ convolved with Object ___h____ produced (d)

iv. Below, write values for a 3x3 filter that when applied to (c) would result in something similar to image (i)

0	0	0
-1	0	1
0	0	0

Vertical edge filter. If lighter side is on the left, the response is negative (dark)

7. True or False (2 pts)

$\text{filter2}(A,\ k \cdot B) - \text{filter2}(A,\ k \cdot C) = k \cdot \text{filter2}(A, B - C)$ (k is a scalar)

True

8. Which of the following is not true about the Fourier transform?
    a. Can be used to improve the speed of linear filtering
    b. Provides a different view of filters that is sometimes easier to interpret
    c. Adds information to the image
    d. Preserves the energy in the image

    C

Templates and image pyramids

Key ideas

- *Implementation and applications of SSD and NCC*
- *Downsampling / upsampling and anti-aliasing*
- *How to compute Gaussian and Laplacian pyramids*

9. Suppose we have computed the normalized cross correlation (NCC), where higher means a better match (max value of 1), and sum of squared differences (SSD), where lower means a better match (min value of 0), for two non-uniform patches. Which of the following are true?
    a. NCC=1 if and only if SSD=0
    b. If SSD=0, then NCC=1
    c. If NCC=1, then SSD=0
    d. None of these are necessarily true

    B

10. Is there any case where NCC is undefined, but SSD can be computed?
    a. No
    b. Yes, if both patches are equal
    c. Yes, if at least one patch has uniform values
    d. Yes, if either patch has negative values

    C

11. Which of these statements are false?
    a. NCC is invariant to shifts in intensity
    b. SSD is usually a better matching cost/score than NCC for calibrated stereo pairs and tracking in video
    c. NCC is faster to compute than SSD
    d. NCC is usually a better matching cost/score than SSD for comparing patches of images from images taken at different times or different cameras

    C

12. If you want to downsample an image by a factor of 2, what is the problem with directly sampling pixels on every other row and column, and how can you fix it?
    a. Aliasing, apply a high pass filter before sampling
    b. Aliasing, apply a low pass filter before sampling
    c. Extrapolation, apply a high pass filter before sampling
    d. Extrapolation, apply a low pass filter before sampling
    e. There is no problem

    B

13. True or False: An image can be losslessly reconstructed from its Laplacian pyramid

    True

Lighting and Color

Key ideas

- *Diffuse vs. specular reflection*
- *Effects of material and geometry on reflection (albedo, surface orientation, viewing direction, etc.)*
- *Use and advantages/disadvantages of color spaces: RGB, Lab, HSV, YCrCb*
- *How to compute histograms*
- *How to improve contrast or balance intensities/color*

14. Under which type of reflection does incoming light scatter from the surface?
    a. Diffuse reflection
    b. Specular reflection

    A

15. Which of the following factors impact the observed intensities of diffusely reflected light? (can choose more than one)
    a. The intensities of the light coming into the reflecting surface
    b. The albedo of the reflecting surface
    c. The orientation of the reflecting surface compared to the direction of incoming light
    d. The orientation of the camera/observer compared to the surface normal

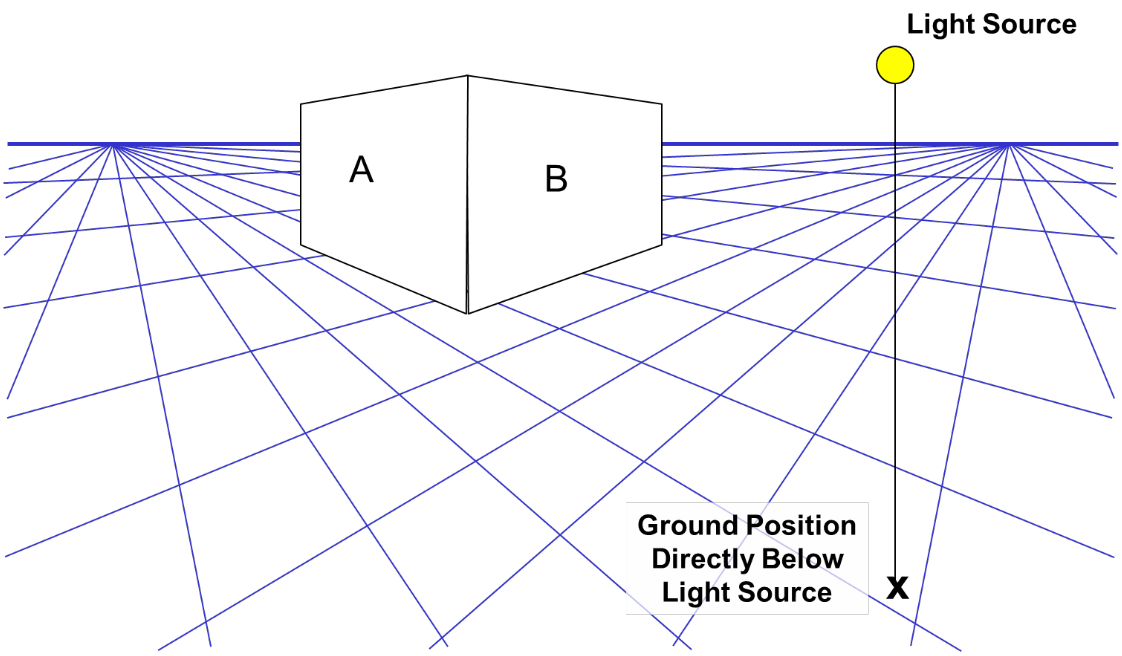
    A, B, C

16. Which of the following factors impact the observed intensities of specularly reflected light? (can choose more than one)
    a. The intensities of the light coming into the reflecting surface
    b. The albedo of the reflecting surface

c. The orientation of the reflecting surface compared to the direction of incoming light
d. The orientation of the camera/observer compared to the surface normal

A, C, D

17. In the illustration below, assume that each side of the depicted block has the same albedo. Which side will look brighter: side A or side B
    a. A will look brighter
    b. B will look brighter



B

18. Imagine that a one-eyed robot is in a dark world, and the only light source is emanating from its eye. Is it possible for the robot to see the shadows that its light source directly creates?
    a. No
    b. Yes

A (only points blocked from line of sight of the light source / eye are in direct shadow)

19. For each material, indicate whether its reflection can be modeled as Lambertian, specular, or mixed:
    a. Balloon Mixed
    b. Cotton shirt Lambertian
    c. Polished wood Mixed

d. Mirror Specular
e. Cement block Lambertian

20. Which color space(s) have separate luminance (brightness) and chrominance (color) channels? (can select more than one)
    a. RGB
    b. HSV
    c. L*a*b
    d. YCbCr

    B, C, D

21. Which color space is based on perceptual studies of just noticeable differences, so that Euclidean distance between nearby color values corresponds to human perception?
    a. RGB
    b. HSV
    c. L*a*b
    d. YCbCr

    C

22. What is a disadvantage of the RGB color space compared to others?
    a. The channels are highly correlated, which makes compression harder
    b. People don't intuitively think of color in terms of red, green, and blue values
    c. Both (a) and (b)
    d. None of these

    C

23. If most of the image intensity values are in the mid-range, such that there are few dark or bright pixels, which of these techniques are best to improve contrast?
    a. Gamma adjustment, gamma > 1
    b. Gamma adjustment, gamma < 1
    c. Histogram equalization

    C

24. If an image looks washed out with many high intensities but not many low, which of these techniques is best to improve contrast?
    a. Gamma adjustment, gamma > 1
    b. Gamma adjustment, gamma < 1
    c. Histogram equalization

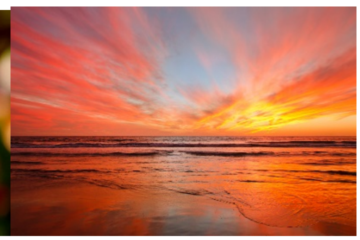
    A (C could also work)

25. If an image is too dark with many low intensities but not many high, which of these techniques is best to improve contrast?
    a. Gamma adjustment, gamma > 1
    b. Gamma adjustment, gamma < 1
    c. Histogram equalization
    B (C could also work)

Pasting: Compositing and Blending

Key ideas

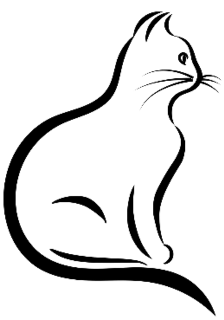
1. *Method and use of alpha matting, poisson blending, and laplacian pyramid blending*

26. Mary wants to cut out the flower region and overlay it on the image of the sunset. She wants to avoid distorting colors or making it look pixelated. What blending method is most appropriate?

a. Use an alpha matte to determine whether the color for each pixel should come from the foreground or background image
b. Use Laplacian pyramid blending
c. Poisson blending (or gradient domain editing)

B




27. Ann wants to see what her arm will look like if she adds this cat drawing as a tattoo. What blending method is most appropriate to overlay the cat on the arm?

a. Use an alpha matte to determine whether the color for each pixel should come from the foreground or background image
b. Use Laplacian pyramid blending
c. Poisson blending (or gradient domain editing)

C

28. Which of the following statements is false?
    a. Poisson blending preserves gradient of the source region without changing the background
    b. Pixels are pretty blocky if you use cut and paste
    c. Foreground colors stay the same when we do Poisson blending
    d. One should feather when doing alpha compositing

C

Image Warping and Morphing

Key ideas

- *Use/conversion of homogeneous vs Cartesian coordinates*
- *Matrix form of transformations: rotation, shear, scale, translation, affine, perspective, similarity*
- *Invariances of transformations*
- *How to solve for transformations from corresponding points*
- *Forward vs. inverse warping/texturing*
- *How to morph shape/texture from one image to another given corresponding points*

29. Which homogeneous coordinate is different from others?

a. (9, 6, 3)
b. (12, 8, 4)
c. (48, 32, 16)
d. (4, 2, 1)

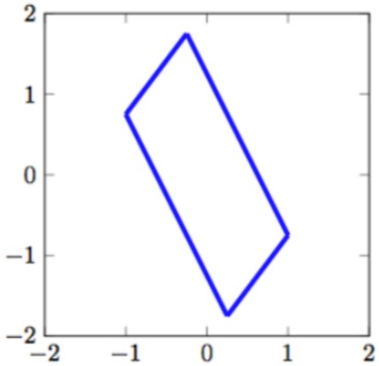
D

30. What properties do affine transformations and projective transformations have in common?

a. Parallel lines remain parallel under projection
b. Origin does not necessarily map to origin
c. Ratios are preserved

B

31. The figure is the output of applying one of the transformations below to a square with vertices (-1, 1), (-1, -1), (1, 1), (1, -1). Which transformation is it (pick most restrictive possible)?



a. Affine
b. Projective
c. Rigid
d. Similarity

A (preserve parallelism)

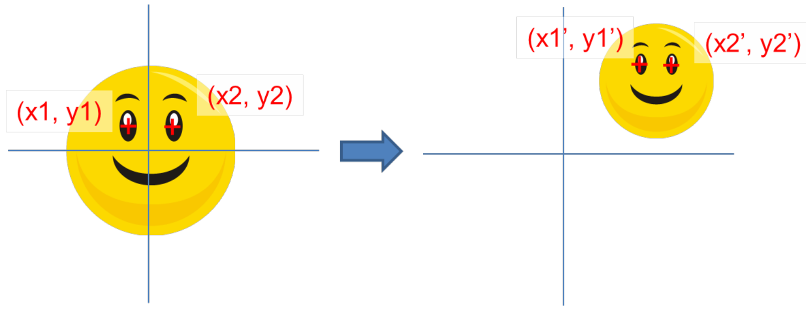
32. What transformations can you model with [a b ; c d]? (can choose multiple)

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

a. Translation
b. Scale
c. Rotation about the Origin
d. Shear
e. Affine
f. Perspective

B, C, D

33. For (i-iii): Suppose that the image of a face is scaled uniformly by factor *s* and translated in either direction by t_x and t_y.



i. Write down the equation for a transformed point (x', y') as a function of the original point (x, y) in terms of t_x, t_y, and s.

$x' = s*x + t_x$

$y' = s*y + t_y$

ii. Write the transformation in matrix form in terms of t_x, t_y, s, x, y, x', and y':

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} & & \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} s & 0 & t_x \\ 0 & s & t_y \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

e. If you are given two pairs of corresponding points: (x1, y1) to (x1', y1') and (x2, y2) to (x2', y2'), how do you solve for the transformation parameters t_x, t_y, s? Write down the system of equations in a matrix form (you don't need to solve it):

$$\begin{bmatrix} x_1 & 1 & 0 \\ y_1 & 0 & 1 \\ x_2 & 1 & 0 \\ y_2 & 0 & 1 \end{bmatrix} \begin{bmatrix} s \\ t_x \\ t_y \end{bmatrix} \cong \begin{bmatrix} x_1{}' \\ y_1' \\ x_2' \\ y_2' \end{bmatrix}$$

34. Put these operations in the correct order to perform morphing: (op1) determine corresponding points; (op2) compute average shape; (op3) define triangulation; (op4) warp both images toward the average shape; (op5) cross dissolve.
    a. (op2) -> (op1) -> (op3) -> (op4) -> (op5)
    b. (op1) -> (op2) -> (op4) -> (op3) -> (op5)
    c. (op1) -> (op3) -> (op2) -> (op4) -> (op5)
    d. (op3) -> (op1) -> (op2) -> (op4) -> (op5)

    C (note: op2 and op3 can be switched)

35. What is the lowest parameter transformation that can map from any point in any one triangle to a corresponding point in any other triangle?

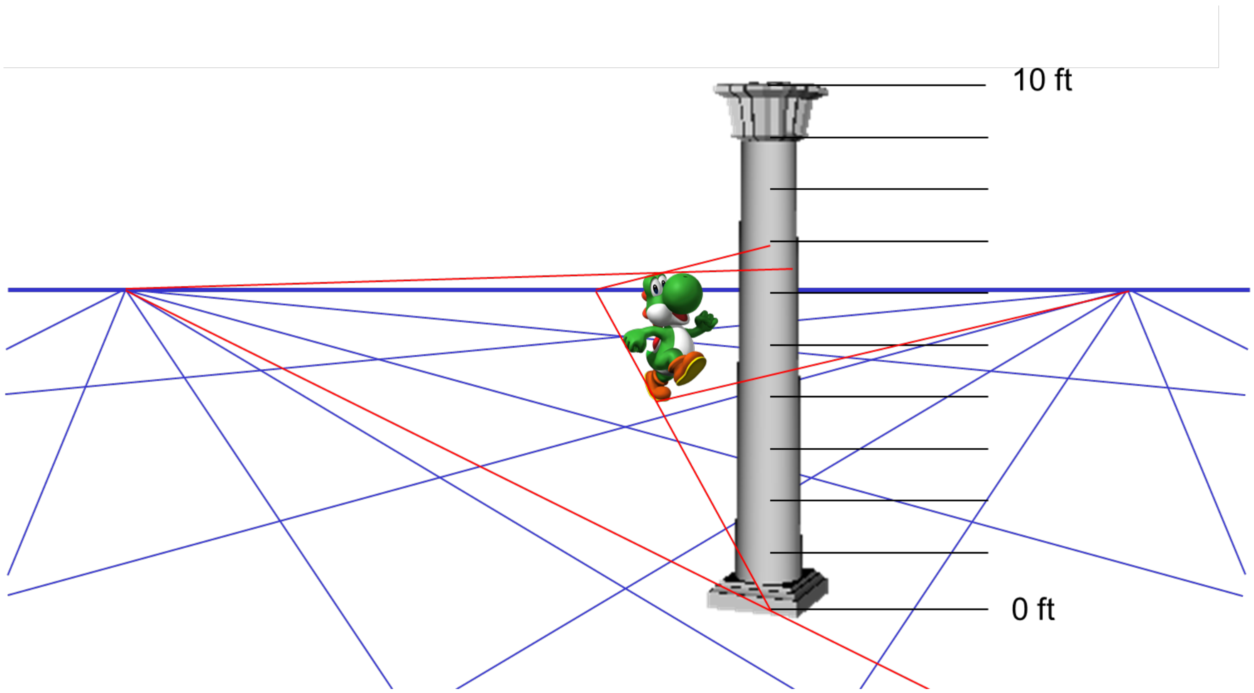
a. Translation
b. Similarity
c. Affine
d. Perspective

C

Single View Metrology and Camera Model

Key ideas

- *Definition of vanishing points/lines and use for comparing object sizes*
- *How to project a 3D point onto the image in a perspective camera model, using equations or matrix operations*
- *Definition and parameterization of intrinsic and extrinsic matrices*
- *Effects of camera parameters such as aperture, lens, focal length, sensor size, image resolution, shutter time*



36. The column is 10 ft tall with the lines evenly marking 1 ft increments, and Yoshi (dinosaur) is on the ground. What is the approximate (within 0.5 ft) height of Yoshi?
a. 4.5 ft
b. 6.0 ft
c. 6.5 ft
d. 7.0 ft
e. None of these are within 0.5 ft

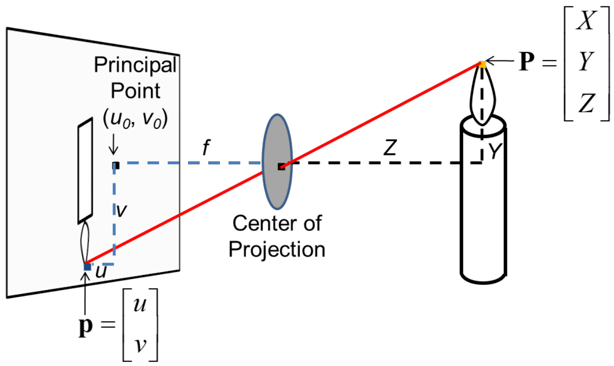
D

37. The column is 10 ft tall with the lines evenly marking 1 ft increments, and Yoshi (dinosaur) is on the ground. What is the approximate (within 0.5 ft) height of the camera?
a. 4.5 ft
b. 6.0 ft
c. 6.5 ft
d. 7.0 ft
e. None of these are within 0.5 ft

B (where horizon intersects with column)

38. Draw a diagram of the projection of a 3D point onto an image pixel and label: vertical component of 3D position *Y*; distance from camera along camera axis *Z*; pixel column *u*; pixel row *v*; principal point u_0, v_0; focal length *f*; center of projection (aka camera center).

See figure below



39. Suppose a point's position is X, Y, Z in the camera's 3D coordinates. Given focal length f and principal point (u0, v0), what is the pixel row *v* of the projected point.

   a. $v = \frac{fY}{X} + v_0$
   b. $v = \frac{X}{fY} + u_0 + v_0$
   c. $v = fY + Zv_0$
   d. $v = \frac{fX}{Z} + u_0$
   e. None of these are correct

E (v = fY/Z+v0)

40. Complete the intrinsic parameter matrix that projects from a 3D point to a 2D homogenous image coordinate. Assume zero-skew and unit aspect ratio.

$$w\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

41. Let (X=2, Y = 0, Z = 10), f=2, u0 = 0.5, v0=0.5. Solve for u.

   a. 0.9
   b. 0.7
   c. 0.4
   d. 0.5

A (fX/Z+u0)

42. Suppose you are taking a picture in low-light conditions and your first photo looks grainy. Which of the following can increase the amount of light coming into the camera without changing the field of view? (can select more than one)
    a. Increase aperture
    b. Decrease aperture
    c. Increase focal length
    d. Decrease focal length
    e. Increase shutter time
    f. Decrease shutter time

    A, E

43. How can you increase the field of view?
    a. Increase aperture
    b. Decrease aperture
    c. Increase focal length
    d. Decrease focal length
    e. Increase shutter time
    f. Decrease shutter time

    D

Single-view 3D Reconstruction

Key ideas

- *How to solve for camera parameters from vanishing points*
- *Single-view scene models, such as single-point perspective stage, ground + billboards*

44. Is it possible to solve for intrinsic matrix using vanishing points that correspond to three orthogonal directions? If so, how?

    Yes, the 3D directions are orthogonal, so e.g. $X_i^T X_j = 0$ where $x_i = KRX_i$. X_i is a 3D direction, x_i is the position of the vanishing point on the image plane, and K is the intrinsic matrix, which includes focal length and principal point (skew, aspect assumed to be 0, 1 respectively). R cancels out in the solution.

45. Suppose you know the camera intrinsics and vanishing line of the ground plane. What additional piece of information could be used to recover the height of the camera?
    a. The height of an object that is visibly connected to the ground
    b. The height of an object that is not visibly connected to the ground
    c. Knowledge of the sensor width
    d. The height of a scene point that projects onto the horizon

    A, D

46. True or False: Single-view 3D reconstruction can work because there is a unique solution to the 3D scene model that results in an observed image of a natural scene.
    False

Image-based Lighting

Key ideas

- *How to create HDR image from multiple LDR images*
- *Equirectangular image representation*
- *Method to create light environment from photos of light probe*
- *Method to relight an object in an image using an environment map*

47. If you want to create a HDR image from multiple images, which of these should vary from image to image?
    a. Shutter time only
    b. Camera position (translation) only
    c. Camera orientation only
    d. Camera position and orientation

    A

Matching, Alignment, Stitching, RANSAC

Key ideas

- *Keypoint detection methods and advantages/disadvantages (DoG, Harris)*
- *Keypoint representations (intensity patch, SIFT)*
- *Keypoint matching criteria (SSD, Lowe's ratio, NCC)*
- *Solving for transformations given matching points (with or without outlier matches), inc. RANSAC and least squares*
- *Perspective and cylindrical image/canvas representations*

(For the two questions below) Suppose we think that two sets of 2D points are related by a homography. We want to estimate the transformation parameters using RANSAC given corresponding pairs of points (which may contain some bad correspondences).

```
For t = 1 to 100:
        Sample N pairs of points
        Solve for the homography transform from sampled pairs
        Get score for this transformation, check if it's the best so far
Return the best scoring transformation
```

48. What (integer) value should N have? _______

 4 (minimum number of points needed to recover transformation

49. If you expect that a large fraction of the correspondences are incorrect (points in a pair don't correspond to the same scene point), what should you do to improve your result?
    a. Increase the number of iterations (e.g. from 100 to 5000)
    b. Increase N, the number of sampled points
    c. Decrease N, the number of sampled points
    d. Instead of RANSAC, compute a least squares estimate of the transformation for all points

    A (you need more iterations to increase probability that at least one sample has no outliers)

50. Suppose you have two sets of corresponding points. The point positions may not be exact, but the correspondences are correct (no outliers). Which method is most suitable to estimate the transformation between the points?
    e. Least squares estimate using all points
    f. RANSAC
    g. Neither can give a good estimate

 A (note: RANSAC would also work, but it's unnecessarily complex since there are no outliers)

51. Which of these is an advantage of the keypoint detection approach used by SIFT, compared to the Harris corner detector? (can select more than one)
    a. Faster to compute
    b. More precisely repeatable detected positions
    c. Provides an estimate of relative scale between two detected points

 C

Object/image recognition and retrieval

Key ideas

- *Techniques to improve speed and accuracy of keypoint-based recognition/retrieval (K-tree clustering, histograms of visual words, inverse index, tf-idf, geometric verification)*

52. Which of these techniques are primarily used for interest-point based image retrieval to improve the **speed** (ignoring effects on accuracy) of matching in large image sets?
    a. K-means clustering
    b. Term frequency – inverse document frequency (tf-idf) score
    c. Index table of which images contain each visual word
    d. Geometric verification of matched keypoints

    A, C

53. Which of the following increase the term-frequency inverse document frequency weight?
    a. The word appears often in the document
    b. The word appears often in the database
    c. The word appears rarely in the document
    d. The word appears rarely in the database

    A, D

Opportunities of Scale

Key ideas

- *Nearest neighbor retrieval as an inference/classification method*
- *Encoder/decoder framework (conceptual)*

54. True or False: With enough data, a simple nearest neighbor lookup can be used to solve complex image labeling and transformation problems.
    True

55. True or False: If you have millions of samples, the features used to represent the samples do not have much impact on the performance of nearest neighbor retrieval
    False

Generating / detecting fakes

Key ideas

- *Understand the techniques behind the fake detection methods (i.e. based on image statistics, mosaicking regularity, JPEG details, lighting direction)*
- *Understand the basic concept of image generation (pix2pix and CycleGAN) – what do they do, what kinds of inputs, what objective or loss are they optimizing*

56. What do we know about the surface normal corresponding to the image boundary of a rounded object?
    a. The normal is parallel to the image plane (Nz = 0)
    b. Nx and Ny are determined by the orientation of the boundary
    c. Neither a nor b

d. Both a and b

D

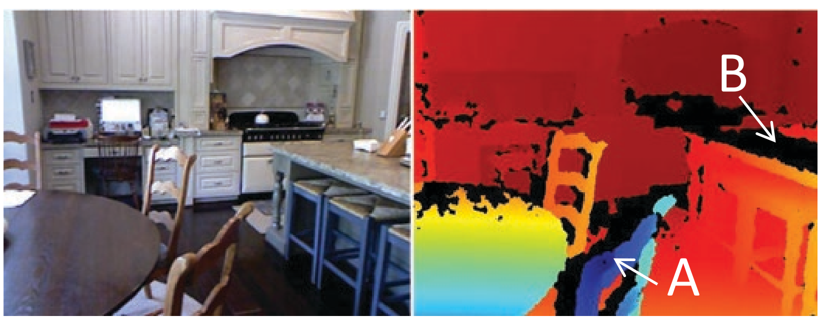
57. Suppose you have an image that was compressed with JPEG quality of 65. Which of these will lead to further degradation of the original image?
    a. Resaving the image with JPEG quality of 80
    b. Resaving the image with JPEG quality of 65
    c. Resaving the image with JPEG quality of 50
    d. a, b, and c
    e. a and c
    f. b and c
    g. none of these

E

Kinect sensor and its use for body pose estimation

Key ideas

1. *Use of stereo/projector pair for stereo depth estimation*
2. *In which cases will stereo likely succeed or fail*



58. Why is the Kinect sensor unable to produce depth values for the countertop, pointed to by "B"? (black means Kinect did not return depth values)
    a. The surface is too slanted
    b. The countertop is shiny (specular reflection), so the projected pattern is not received or matched by the camera
    c. The surface doesn't have enough texture
    d. The area of the surface in the image is too small

B

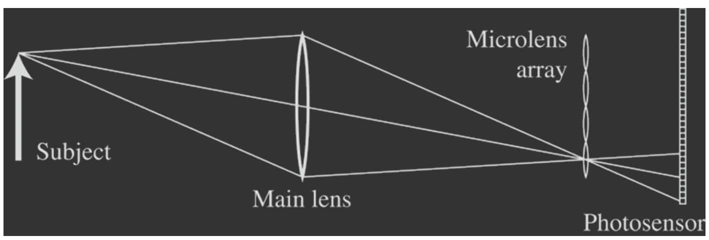
59. The Kinect stereo algorithm uses NCC, rather than SSD for a match score. Is NCC clearly a better choice than SSD in this case. If so, why?
    a. No, either SSD or NCC could be effective in this case
    b. Yes, the projected image is known so we know what intensities the sensed image should match
    c. Yes, the sensed image will be locally proportional to the projected image, but the overall intensity could be different

C

Computational approaches to cameras

Key ideas

1. *Why the coded aperture enables better depth estimation and deblurring*
2. *How the plenoptic camera works to capture both the position and angle of rays coming into camera*



60. Circle which camera has the advantage in each of the following ways. (3 pts)
    C = conventional, **P** = plenoptic (shown above), **N** = neither

    i. Can produce a higher resolution image C P N

    ii. Captures more light C P N

    iii. Encodes images with multiple camera centers C P N

C, N, P

Face detection, recognition, transformation

Key ideas

- *Basic idea of classic face detection (image pyramid, sliding window, extract features, apply trained classifier), and what some of the challenges are*
- *Face recognition via detection, alignment, feature extraction, and similarity comparison; what parts of the modern face recognizer provides the most improvement*
- *How to make a face look more masculine, happy, young, etc. through warping, averaging, and interpolation*

61. Which of these is **not** a step in the warping/transformation method of making an input young face look older? (can choose more than one)
    - **a.** Obtain an average “young” face, including triangulated feature points
    - **b.** Obtain an average “old” face, including corresponding feature points
    - **c.** Define corresponding feature points on the input face
    - **d.** Define a new triangulation on the input face using the Delaunay algorithm
    - **e.** Create a target “shape” as Xtarget = Xinput + alpha*(Xold – Xyoung), where alpha is a deformation factor greater than 0
    - **f.** Obtain the affine transformations for each triangle in Xtraget to Xinput, Xold, and Xyoung
    - **g.** For each pixel in the target image, map to the input, young, and old faces and store the color as Ctarget = Cinput + alpha *(Cold-Cyoung)
    - **h.** Apply a sharpening filter to accentuate wrinkles

    D, H (the same triangulation, which is a mapping of vertices to faces, is used by all faces)

Video magnification

Key ideas

- *Basic workings and advantages/disadvantages of the point tracking approach, the pixelwise intensity magnification, and the phase magnification approaches to motion magnification*

62. Which of these is an advantage of the point tracking method for magnification, compared to the Eulerian and Phase-based?
    - a. Point tracking method is simpler and more robust
    - b. Point tracking method usually generates more compelling results
    - c. Point tracking method can be applied to large non-periodic motions

    C

63. What are advantages of the phase magnification approach, compared to the pixel magnification approach?
    - a. Performs translation more directly and so can apply to larger movements than intensity magnification
    - b. Works well with non-period motions
    - c. Robust to imaging noise

    A, C